(12) UK Patent Application (19) GB (11) 2 178 569 A

(43) Application published 11 Feb 1987

(21) Application No 8525682

(22) Date of filing 18 Oct 1985

(30) Priority data
(31) 8519004 (32) 27 Jul 1985 (33) GB

(51) INT CL⁴
G11B 21/02 G06F 12/00 G11B 25/04

(52) Domestic classification (Edition I):
G4A MG2

(56) Documents cited
GB A 2100024

(58) Field of search
G4A
G3N
Selected US specifications from IPC sub-classes G06F G11B

(71) Applicant
Miro Jan Germuska,
47 Clifden Road, Worminghall, Aylesbury, Bucks HP18 9JR

(72) Inventor
Miro Jan Germuska

(74) Agent and/or Address for Service
Miro Jan Germuska,
47 Clifden Road, Worminghall, Aylesbury, Bucks HP18 9JR

(54) Relational database storage on magnetic and optical disks

(57) A disk drive for storing data in a tabular format provides fast access to both rows and columns by superposing a pair (5, 7 Fig. 4) of oscillators upon an overall spiral motion (4) of a head able to access records on the disc, so that segments in a plurality of different tracks may be accessed within one revolution of the disc. Two heads oscillating in antiphase may enable all segments in all tracks to be so accessed (Fig. 5). The data may be stored in such a way that when an element changes, only that element is recorded again and not the record to which it belongs. Write-once-only optical disks may thus be made suitable for storage of live data and not just for archiving.

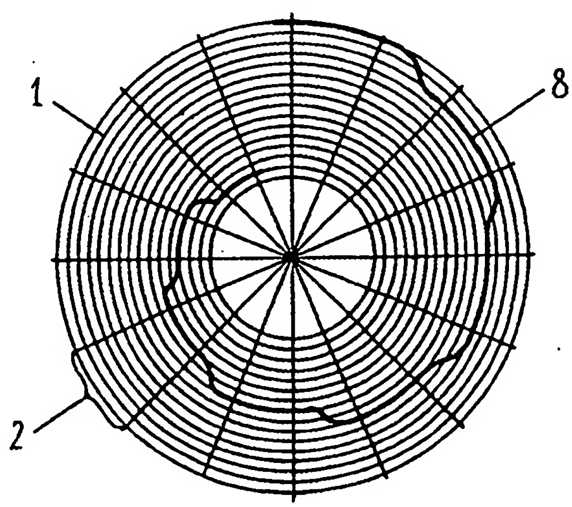

FIG 3

GB 2 178 569 A

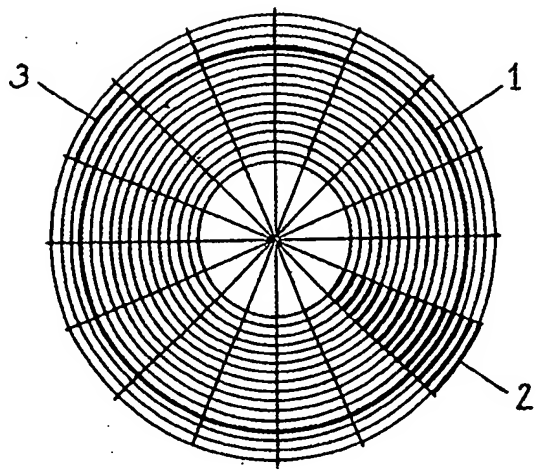

FIG 1

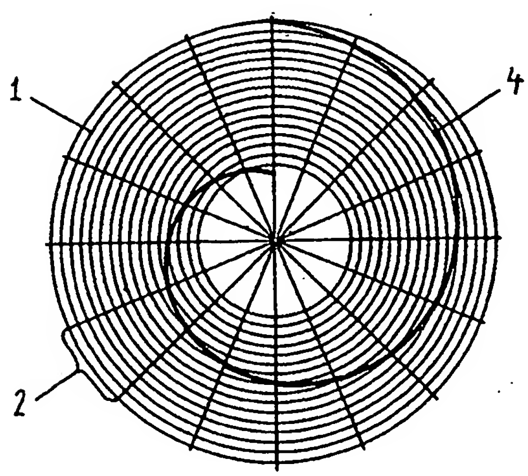

FIG 2

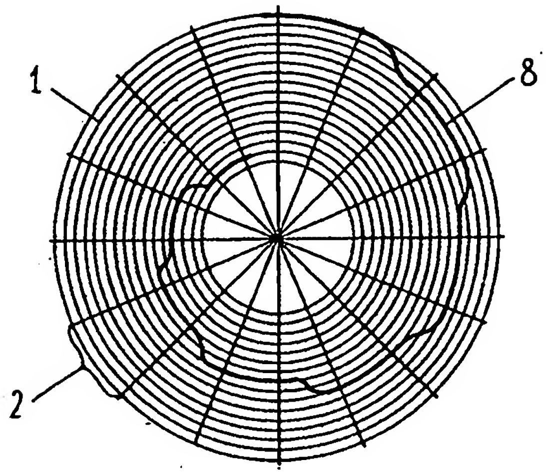

FIG.3

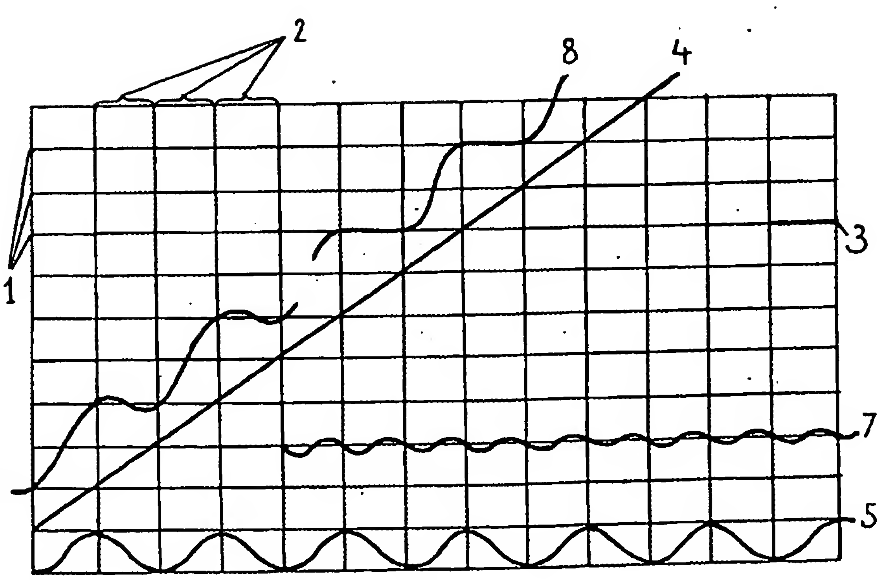

FIG 4

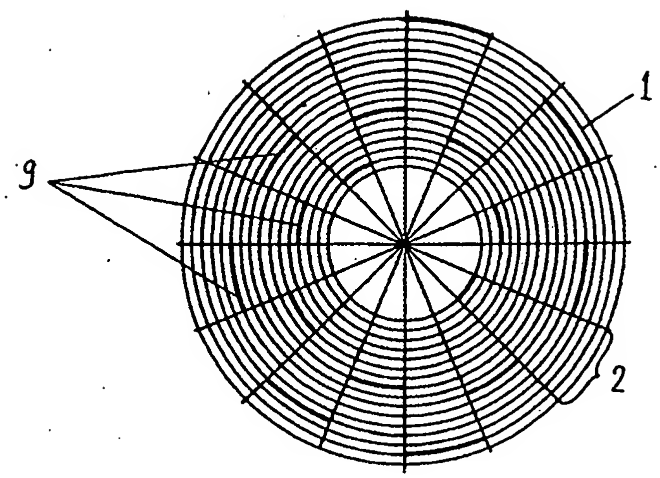

FIG 5

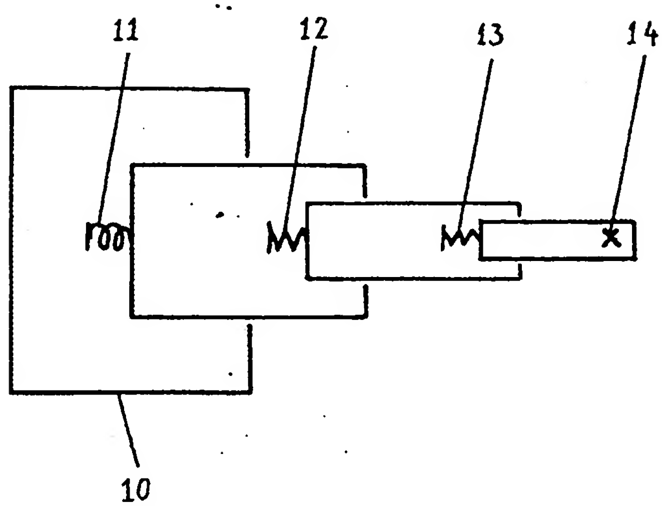

FIG 6

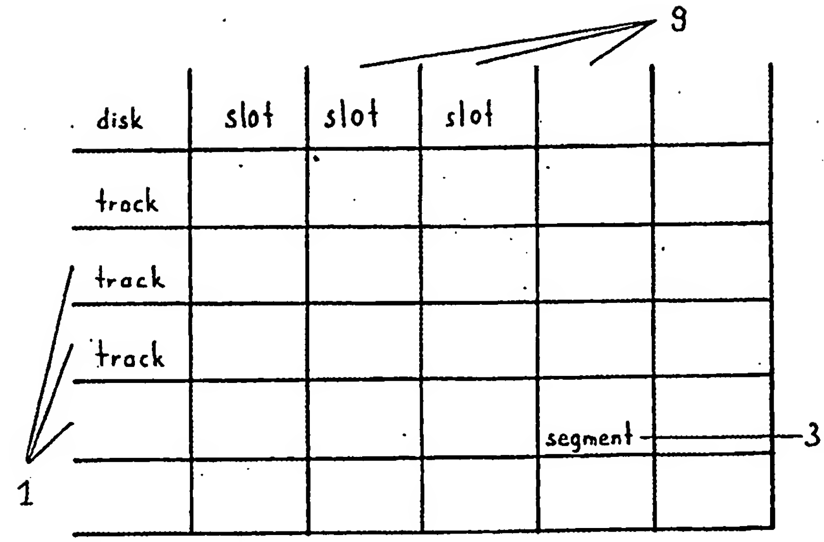
9
disk
slot
slot
slot
track
track
track
segment
3
1

FIG 7

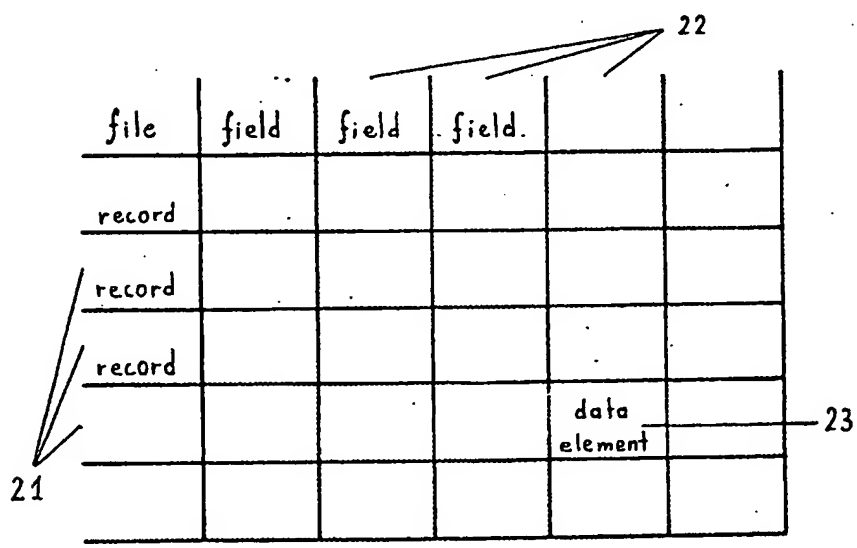
22
file
field
field
field.
record
record
record
data element
23
21

FIG 8

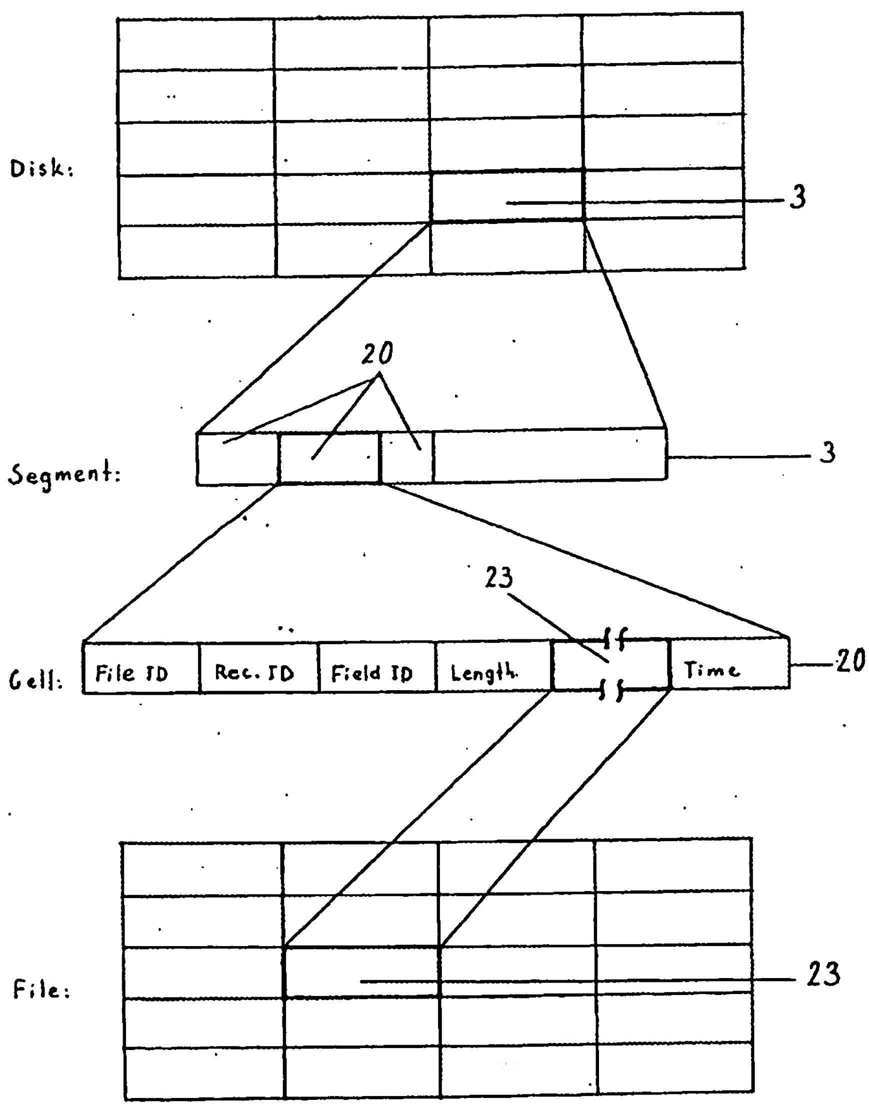
Disk:
3
20
Segment:
3
23
Cell:
File ID
Rec. ID
Field ID
Length
Time
20
File:
23

FIG 9

SPECIFICATION

Relational database storage on magnetic and optical disks

This invention relates to the design of computer disk drives suitable for use by relational databases. Also to a method of storing computer data in need of perpetual updating on optical disks.

Conventional disk drives record data on the disk in concentric tracks in such a way that in one revolution one can read/write one track of data. One track may contain several records and one record several data elements. In order to locate required records quickly various indexes to physical positions of records are maintained.

Records can by physically grouped together with respect to one primary index. With respect to secondary indexes the records are scattered throughout the disk at random. Thus accessing data via secondary indexes in general requires a separate disk revolution for each record involved. This of course is very time consuming. Optical disks have a further disadvantage in that the laser burns are irreversible and every time a data element changes the entire record has to be rewritten to a fresh space. Thus, although their capacity is at least 10 times greater than for comparable magnetic disks their use is restricted to archiving.

The present invention seeks to overcome the above drawbacks. It does so by providing a method of storing individual data elements in a tabular format and by providing a mechanism for reading/writing either a row or a column of data in one disk revolution.

The invention will now be described in greater detail, by way of example, with reference to the following drawings:

Figure 1 shows a plan view of the disk for explaining the positions of tracks, sectors and segments.

Figure 2 shows the spiral trajectory relative to the disk that the read/write head would follow in the absence of any oscillation.

Figure 3 shows the required head trajectory relative to the disk that the read/write head should follow.

Figure 4 shows diagrammatically how the linear movement is combined with the oscillation in order to produce the required head movement.

Figure 5 shows a possible slot pattern consisting of one segment from each track.

Figure 6 shows diagrammatically the structure and the components of the read/write mechanism.

Figure 7 shows the logical structure of the disc divided into tracks and slots.

Figure 8 shows the logical structure of a flat data file divided into records and fields.

Figure 9 shows the relationship between the disk and a file and also what is recorded for each data element.

The data on the disk is recorded in concentric tracks 1 which are divided into segments 3 forming angular sectors 2 as shown in Fig. 1. Alternatively the tracks form one continuous spiral, not shown on any drawings. The number of tracks is shown to be the same as the number of sectors. This arrangement will be considered first as it is easiest to explain. Later it will be shown how a larger number of tracks may be used. Also in practice the number of sectors will be much larger than shown, for example 256 or more.

In order to access a multiplicity of segments from one track the arm is moved so that the read/write head is positioned over the track. The head waits in that position and is activated when the required segments pass underneath it to read or write as needed. If the tracks are on a continuous spiral the head moves at a constant velocity following the spiral over the required track.

In order to access segments from a multiplicity of tracks the arm is moved radially across the tracks at a constant velocity so that the head follows a spiral 4 relative to the rotating disk as shown in Fig. 2. At the same time the head oscillates so that it follows a stepwise trajectory 8 as shown in Fig. 3 and Fig. 4. If there are two arms moving in unison while their heads oscillate half a cycle phase difference then when one head is stationary over a track the other is moving across the tracks. If the heads are separated by an even number of sectors then the two heads between them access one segment from each track in one disk revolution. The set of segments accessed in this way will be referred to as a slot 9, an example of which is shown in Fig. 5.

Referring back to Fig. 4 the arm movement is represented by line 4 and the oscillation by line 5. The amplitude of the oscillation is equal to the distance between the centres of neighbouring tracks 1 and the frequency is such that one cycle is completed while the disk rotates by two sectors 2. Assuming sinusoidal oscillation the deviation of the head trajectory 6 from the track centre is greatest at approximately $\frac{1}{4}$ and $\frac{3}{4}$ through each cycle. At those points it is about $^1/_{10}$ of the amplitude. In order to eliminate this deviation a secondary oscillation within the amplitude 10 times smaller and the frequency 2 times greater as shown by the line 7 may be superimposed on the primary oscillation to produce the head trajectory 8 virtually parallel with the track.

It will be appreciated that it may be difficult to maintain the constant velocity of the arm let alone to synchronise its movement and the head oscillation with the rotation of the disk. This problem will be addressed next. Referring to Fig. 6 the entire arm 10 is accelerated by a strong stepper motor which is not shown or by some other accelerator. The accelerator is

switched on and off according to the disk position. The primary oscillator 12 and the secondary oscillator 13 are driven by signals produced by the rotating disk so that the oscillation is synchronised with the sector boundaries.

On alternate sectors a head 13 follows a line parallel to the tracks. When the accelerator is switched off this line may not coincide with the centre of the track due to the deviation in performance of the accelerator. However, as long as the performance of the accelerator is consistent enough to leave the arm on a spiral passing through the required segments this deviation can be corrected.

The conventional head tracking methods may be used to determine the head deviation from the centre of the track. The tracking apparatus is switched on by the position of the disk when the accelerator is switched off or shortly thereafter and is switched from head to head on each sector boundary. The signal produced by the tracking apparatus is used to control the voice-coil 11 which provides the driving force to compensate for the deviation in performance of the stepper motor and for the friction.

Each successive part of the arm is several times lighter than the previous one so that the recoil effect caused by activating a motor linking adjacent parts is kept small. Small deviations of the head from the track centre due to the recoil is detected by the tracking apparatus and corrected by the voice-coil 11. This arrangement is shown in order to explain the driving forces involved. In practice it may be possible to use a smaller number of motors to provide the required combined movement. For example, one motor may be used to move the arm and another to provide the oscillation.

So far only the basic arrangement where the number of tracks is equal to the number of sectors have been considered. This basic arrangement may be considered to relate to one band of tracks on the disk. In order to increase the number of tracks several bands may follow immediately each other. If the arm continues moving with the constant velocity it will access the next band on each subsequent disk rotation. Alternatively there may be a pair of heads and a separate buffer for each band of tracks so that all data can be accessed in one disk revolution.

Logically, a disk can be regarded as a two dimensional table shown in Fig. 7 where the rows represent the tracks and the columns represent the slots. Within a computer system data is normally organised into files which contain records which in turn consist of fields. The value of a field in a record is called a data element. We will concern ourselves with flat files as shown in Fig. 8 where each record 21 within a file consists of the same fields 22, although some data elements 23 may be absent or null.

Referring to Fig. 9 a data element 23 is placed into a segment 3 by means of a hashing algorithm. The file identifier and the record identifier determine the track while the file identifier and the field identifier determine the slot. Each element is thus assigned a unique home segment. The converse, however, is not true and one segment may be the home for a multiplicity of elements. Consequently the elements cannot be recorded on their own, but each is recorded within a cell 20 accompanied by its file, record and field identifiers. In order to be able to reconstruct a record as it was at any given time the date and time when the element was written is also recorded. To allow for variable length elements the element length is recorded too.

A record is read from the disk as follows. The track where the record is situtated is calculated by the hashing algorithm from the file and the record identifiers. The entire track is then read into one or a multiplicity of filter buffers. Each filter buffer has its own processor which scans through looking for the elements belonging to the given record and puts them into the I/O buffer. The I/O buffer is then made available to the application requesting the record.

A record is written to the disk as follows. The new elements are transferred to the I/O buffer in the format as above from the application wishing to write the record. The track is calculated and transferred into the filter buffers as for reading. Then the individual elements are transferred from the I/O buffer to the filter buffers where they are appended to the end of data in their respective home segments. Finally, the filter buffers are copied back to the track.

In the case of write-once-only optical disks it is required not to touch a disk area by the laser more than once. For that purpose the old data in the filter buffers is erased before the buffers are copied back to the disk. This may be achieved, for example, by temporarily recording in the filter buffers in a reserved area at the beginning of a segment the length of the data before any new data is appended there.

Often records are needed not on the basis of their identifiers, but because of their contents. For example one may require all orders for a given customer. Traditionally this is accomplished by the use of secondary indexes. With a disk drive that can access a slot one can read/write all elements belonging to a field in a similar way that one can read/write all elements belonging to a record. The slot where the field is situated is calculated by the hashing algorithm from the file and the field identifiers and the reading/writing then proceeds exactly as for records.

In addition, the elements that are transferred to the I/O buffer may be restricted to only those that fall within a specified range of

values. This way one can very quickly identify for example all orders for a given customer, all orders over a certain value, etc.

The hashing algorithm distributes the elements more or less at random amongst the segments and some segments get full before others. To enable the disk to accept data when a required segment is full some overflow mechanism has to be provided. When an element does not fit into its home segment it will be recorded into the next segment on the same track or the next segment in the same slot. If there is no room in either of the two segments the element will be recorded in the segment on the next track in the next slot. After, that further tracks and slots may be used in a similar fashion.

When an element has to go into an overflow its record identifier is recorded in the first segment of the track which is reserved for that purpose. Using this information the disk system can then determine if the next track has also to be read in order to read a given record. Similarly the first segment of each slot is reserved for the overflow field identifiers. Using this information the disk system can determine if the next slot has also to be read in order to read a given field.

The file, record and field identifiers are usually meaningful strings of characters. As such they are quite lengthy and a lot of space on the disk can be saved if they are replaced by meaningless sequential numbers. Thus each file can be given a number. Within a file each field and each record can also be numbered. The names may be treated as any other elements and recorded on the disk. The first record of each file may consist of the field names and the first field of each record may contain the record name. The file name may be recorded in the first field of the first record.

CLAIMS

1. According to the first aspect of the invention a disk drive comprising a rotating disk and a read/write mechanism incorporating a head able to access data from a large number —tens or hundreds—of tracks in one disk revolution.

2. Further, according to the first aspect of the invention a disk drive comprising a rotating disk and a read/write mechanism a part of which moves in a continuous spiral relative to the disk surface thus enabling access to data from a multiplicity of tracks in one disk revolution.

3. Again, according to the first aspect of the invention a disk drive comprising a rotating disk and a read/write mechanism a part of which periodically oscillates thus enabling access to data from a multiplicity of tracks in one disk revolution.

4. According to the second aspect of the invention a method for storing tabular data on a rotating disk where a row is mapped onto a set of segments all of which come from the same track and a column is mapped into a set of segments none of which come from the same track.

5. Further, according to the second aspect of the invention a method for storing data on a rotating disk in such a way that all data elements from one record are stored independently from one another so that when one of them changes no other needs rewriting.

6. Again, according to the second aspect of the invention a method for storing data on a rotating disk in such a way that with the exception of file, record and field identifiers all data elements are stored only once and any record can be reconstructed as it was at any time with the accuracy to which the time is recorded.

Printed for Her Majesty's Stationery Office
by Burgess & Son (Abingdon) Ltd, Dd 8817356, 1987.
Published at The Patent Office, 25 Southampton Buildings,
London, WC2A 1AY, from which copies may be obtained.